Opportunities in data curation: Integrating the library into the research process

Heather Coates, IUPUI University Library

Electronic Resources & Libraries 2012 | April 4, 2012

Goals for today

- Increase awareness of the federal policies encouraging greater access to federally-funded data.
- Identify new roles and activities for librarians in data curation.
- Identify at what phases of the research process the library can integrate data and other services.

Overview

1. Federal data policies
2. The research process
   - @ IUPUI
   - @ your institution
3. Identifying new roles
   - Needs & opportunities
   - Existing expertise
4. Building partnerships

FEDERAL DATA POLICIES…OH MY!

Federal data policies

- National Research Council (1985): Sharing Research Data
- Office of Management & Budget (1999): Circular A-110
- NSF: data sharing policy, DMP requirement (2011)
- NIH : data sharing policy (2003), public access policy (2008)
- NEH, Office of Digital Humanities: Data Management Plans
- CDC Policy on Releasing & Sharing Data: http://www.cdc.gov/od/foia/policies/sharing.htm
- Council on Governmental Relations: Access to and Retention of Research Data: Rights and Responsibilities, (2006)
- National Institute of Justice: Data Resources Program
- Howard Hughes Medical Institute: Research Policies
- NOAA: Proposed Data Sharing Policy

NIH data sharing policy

- $500,000 or more in direct costs in any year of the proposed research
- Final research data, not summary statistics or tables, not underlying pathology reports and other clinical source documents, might include both raw data and derived variables
- If an application describes a data-sharing plan, NIH expects that plan to be enacted.
- NIH expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset.
- It is the responsibility of the investigators, their Institutional Review Board (IRB), and their institution to protect the rights of subjects and the confidentiality of the data. Prior to sharing, data should be redacted to strip all identifiers, and effective strategies should be adopted to minimize risks of unauthorized disclosure of personal identifiers.

NSF data policies

- Driver – greater impact of research dollars
- Encouraging two disparate sets of activities
  - Data management & preservation
  - Data sharing
- Scholarly impact
- Planning ahead to provide for preservation, persistent access
  - If you can't find it, it doesn't exist

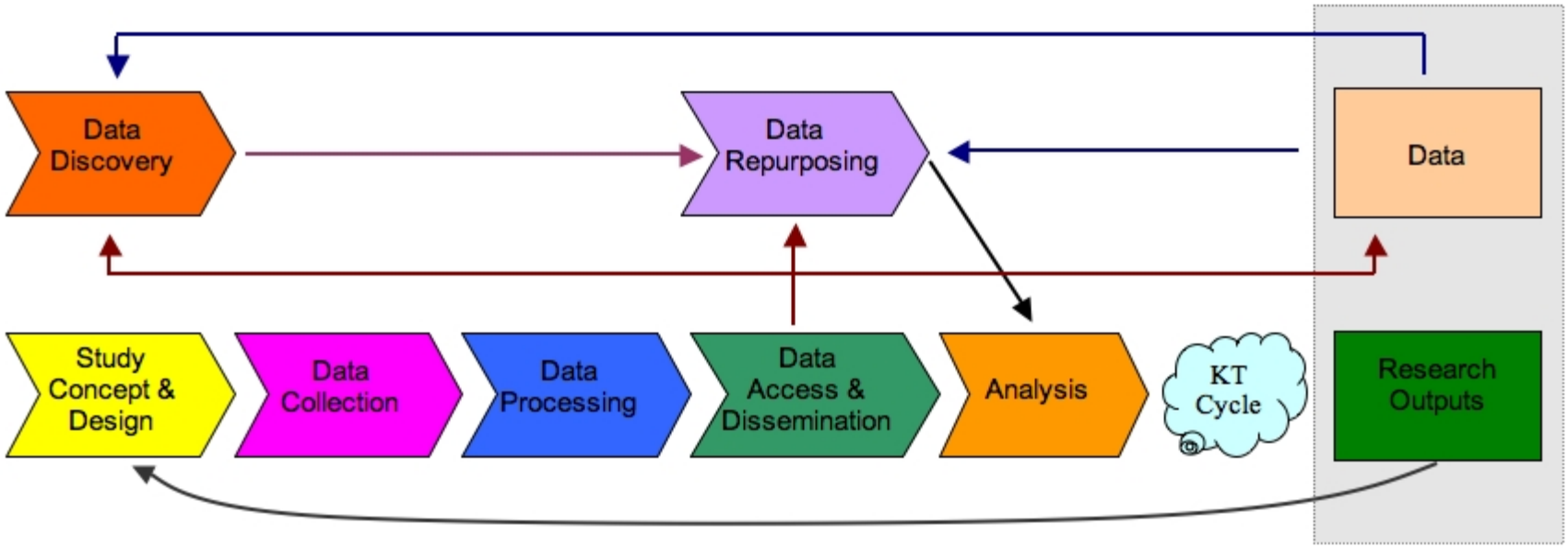
Other data policies: Common themes

Reflect best practices in [your] field

- Data management planning
  - Begins at project conception
  - Planning occurs throughout the data life cycle
  - Important for scientific record, data integrity
- Data sharing
  - Balance privacy, security, intellectual property issues with advancing scientific knowledge
- Data preservation
  - Determine what is valuable (who, how) and how long it will likely be valuable
  - Identify data stewardship roles

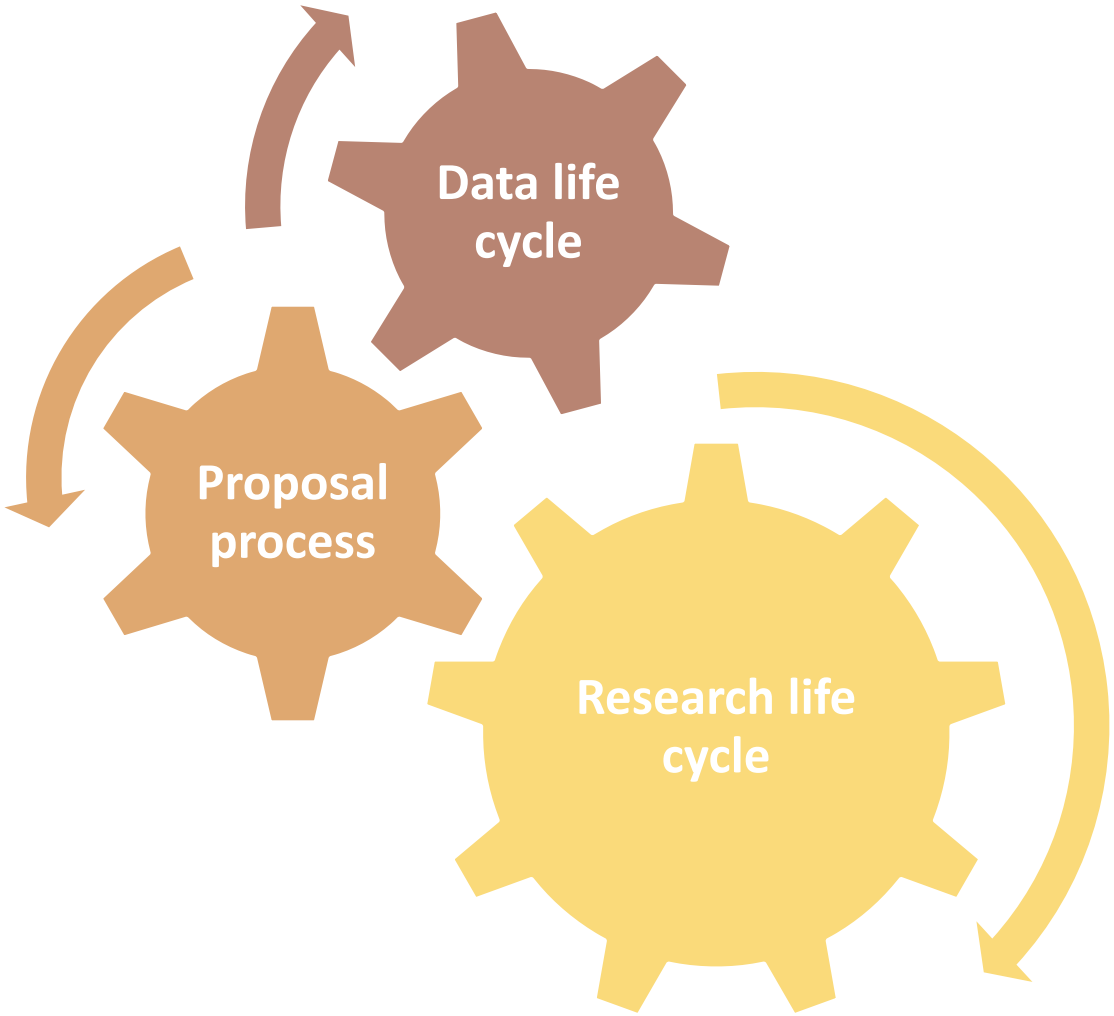
THE RESEARCH PROCESS

The research process

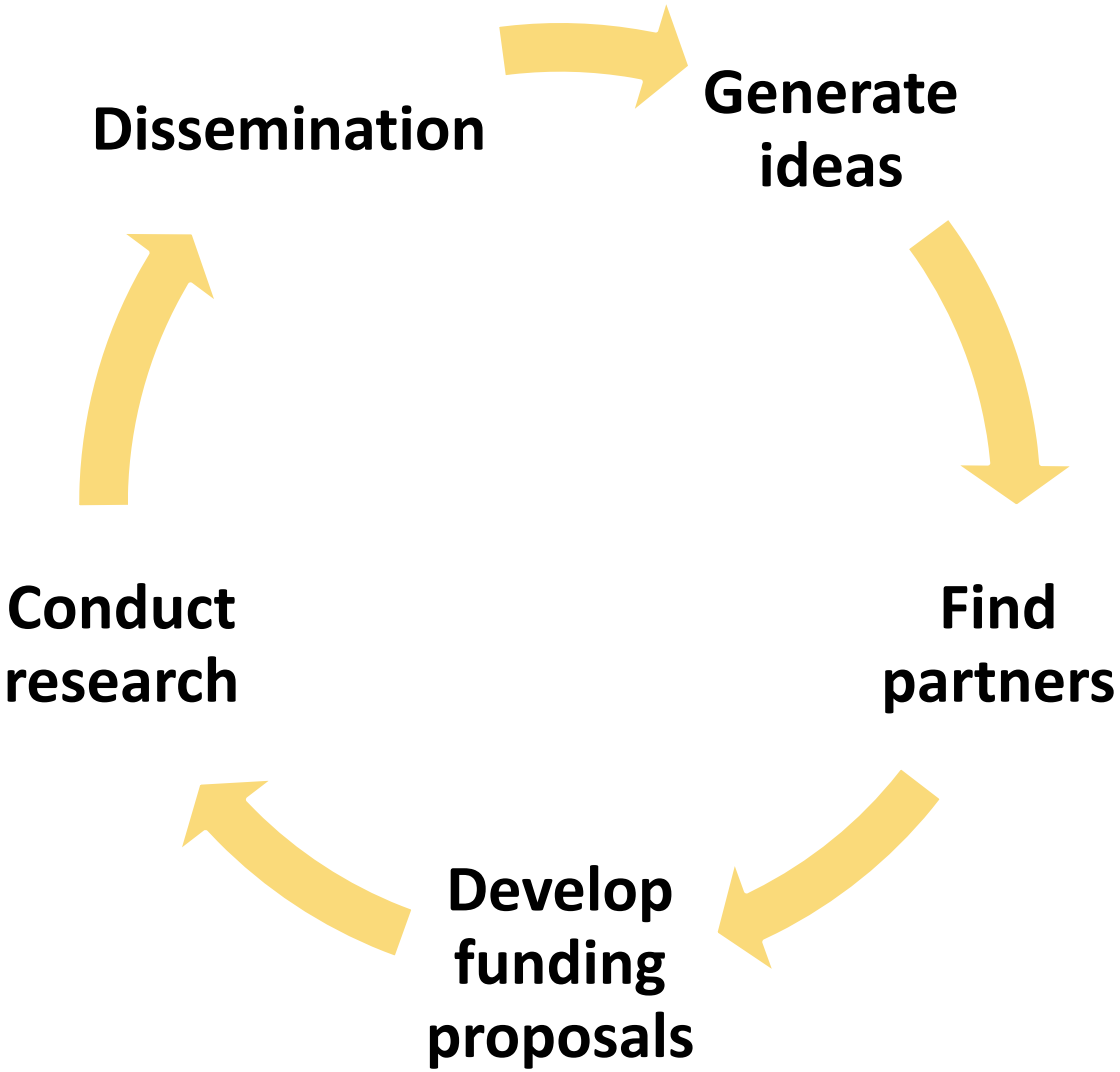


Humphrey, C. (2006). E-Science and the Life Cycle Model of Research. Retrieved from http://datalib.library.ualberta.ca/~**humphrey/lifecycle**-science060308.doc.

Doing research in a digital age

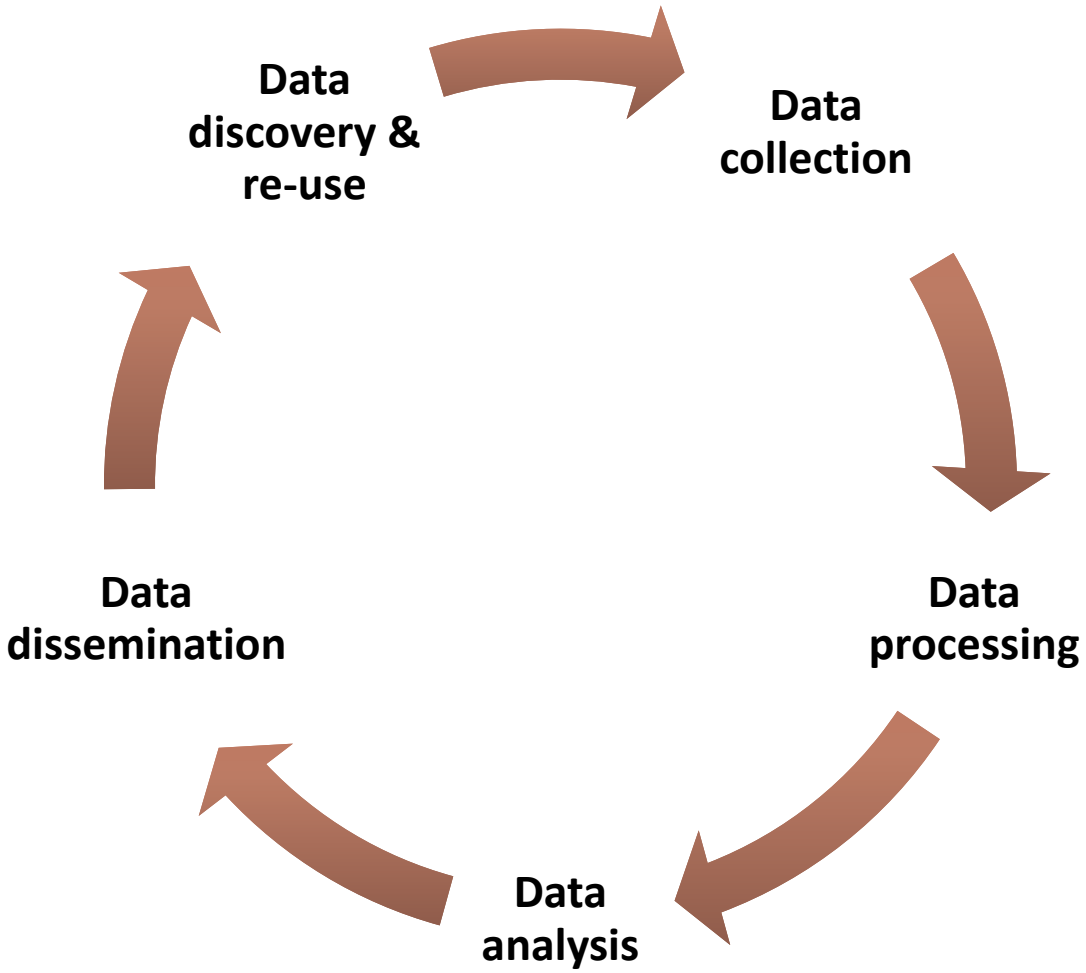


Research Life Cycle



Proposal development process

Data life cycle



Research partners @ IUPUI

- Office of the Vice Chancellor for Research (OVCR)
- Office of Research Administration (ORA)
  - Contracts & Grants
  - Compliance: IRB, IACUC, Chemical Safety, etc.
- Council of Associate Deans for Research (CAD)
- University Information Technology Services (UITS)
  - Pervasive Technology Institute (PTI)
  - Data to Insight Center (D2I)
  - Research Technologies Center (RTC)
- Clinical Translational Sciences Institute (CTSI)
- University Library (UL)

Research @ your institution

- OVCR/OVPR
- Research Administration
  - Contracts & Grants
  - Compliance: Human Subjects, Animal Care & Use, etc.
- Associate Deans for Research
- IT
- Library(ies)

IDENTIFYING NEW ROLES

Why the library?

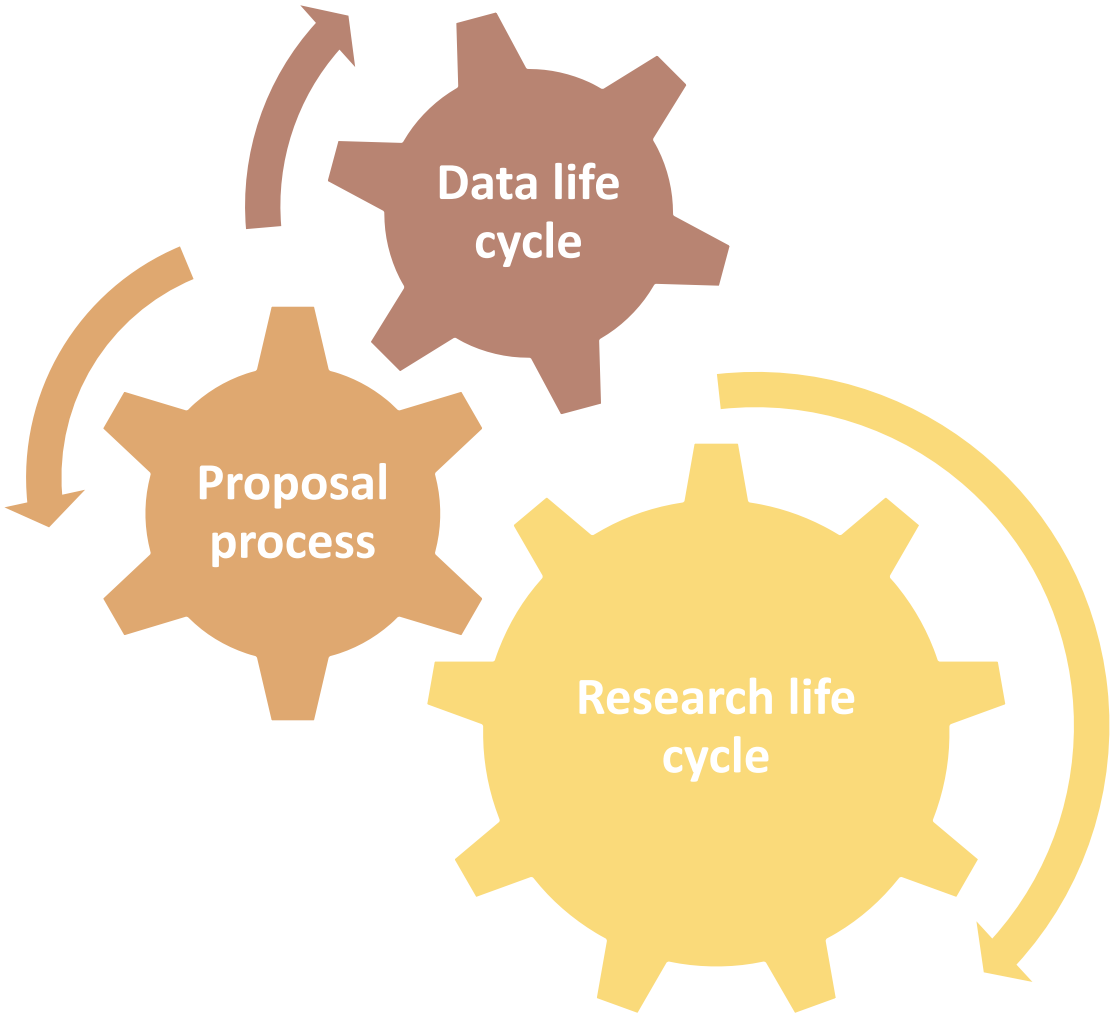
- Libraries are trusted members of the institution
- Libraries are interdisciplinary by nature
- Our organizational structure lends itself to collaboration
- Existing infrastructure for digital information
- Existing expertise in preserving and providing access to information, including data

access

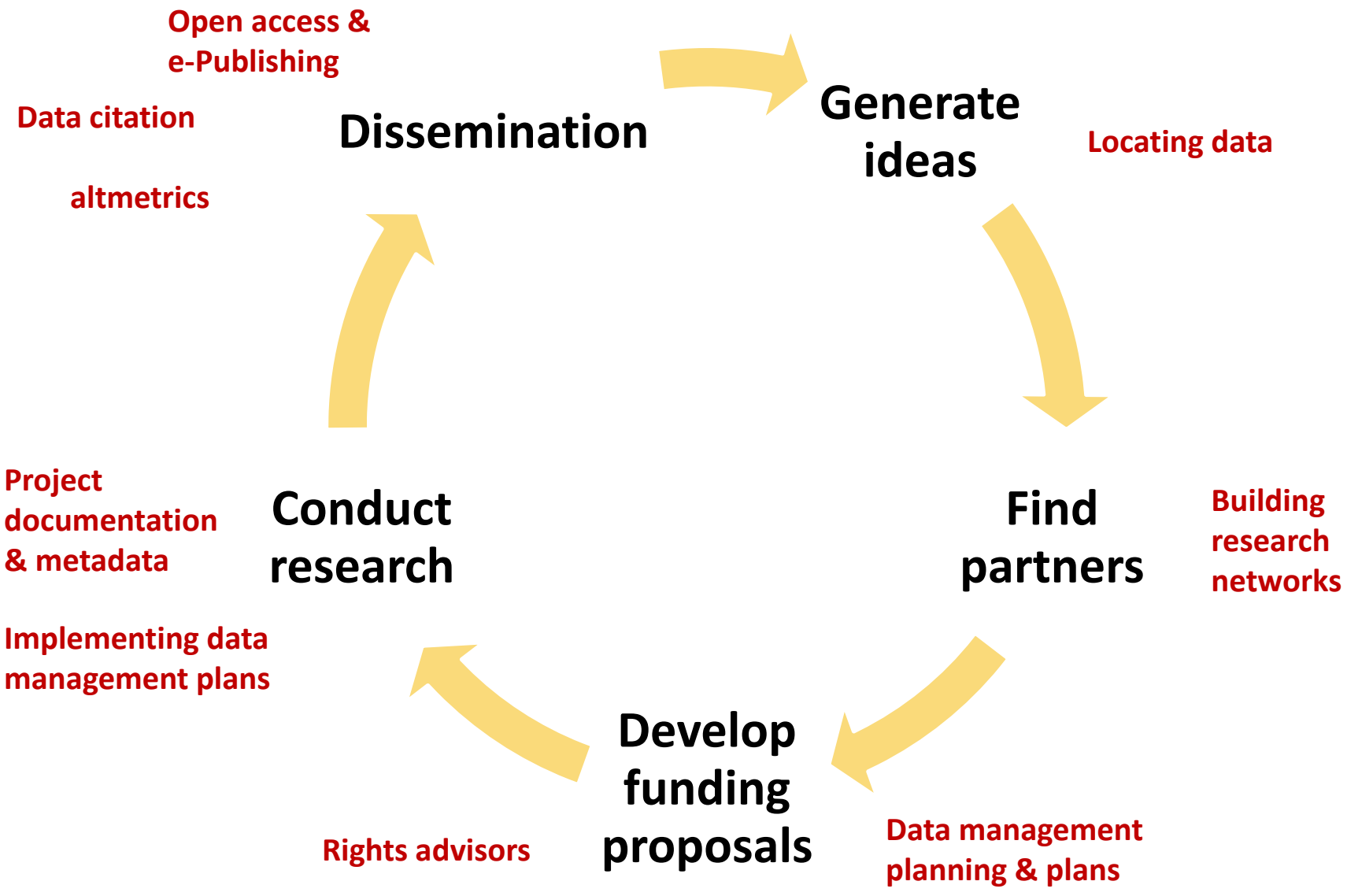
preservation

curation

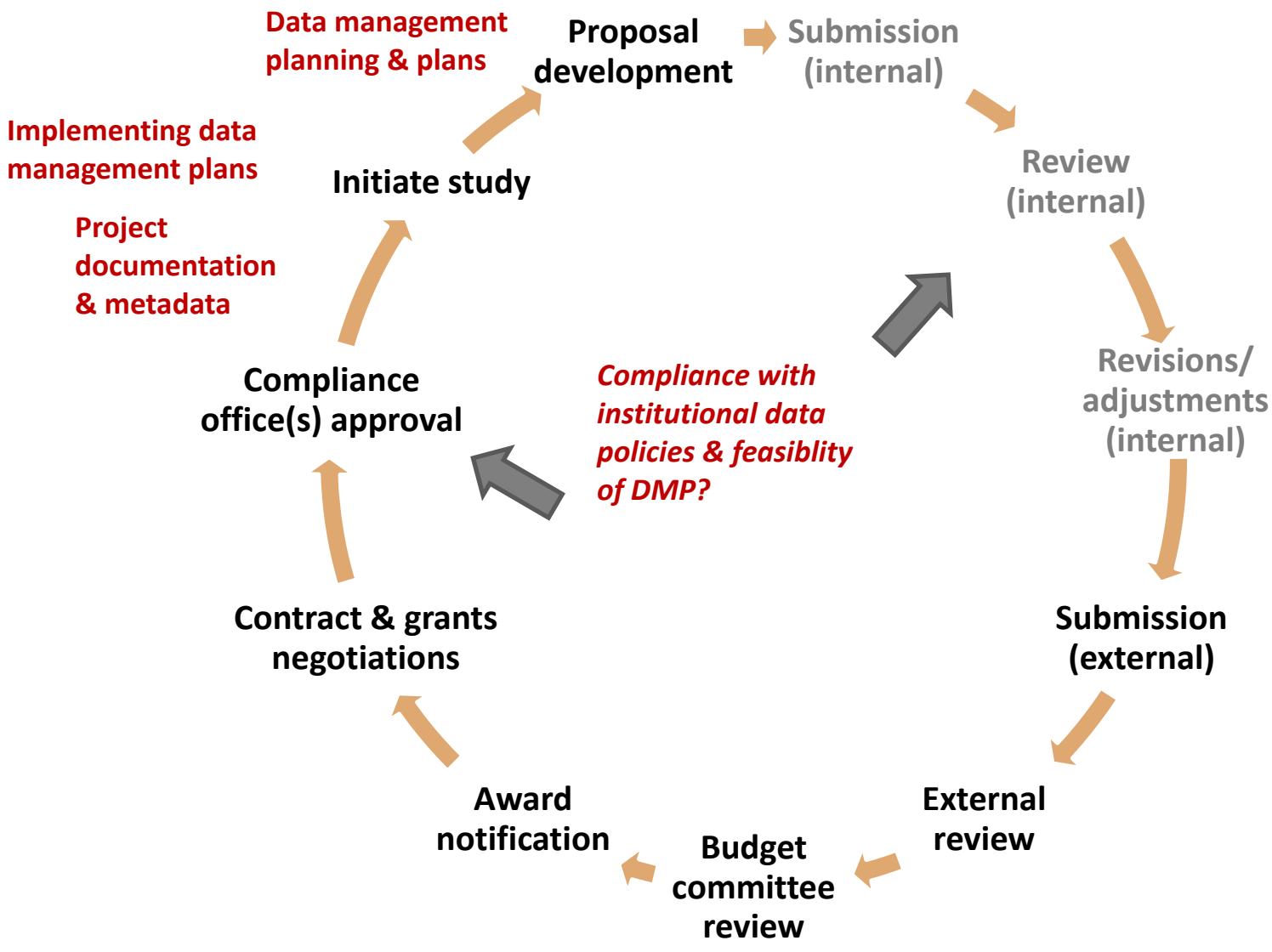
Where does the library fit?



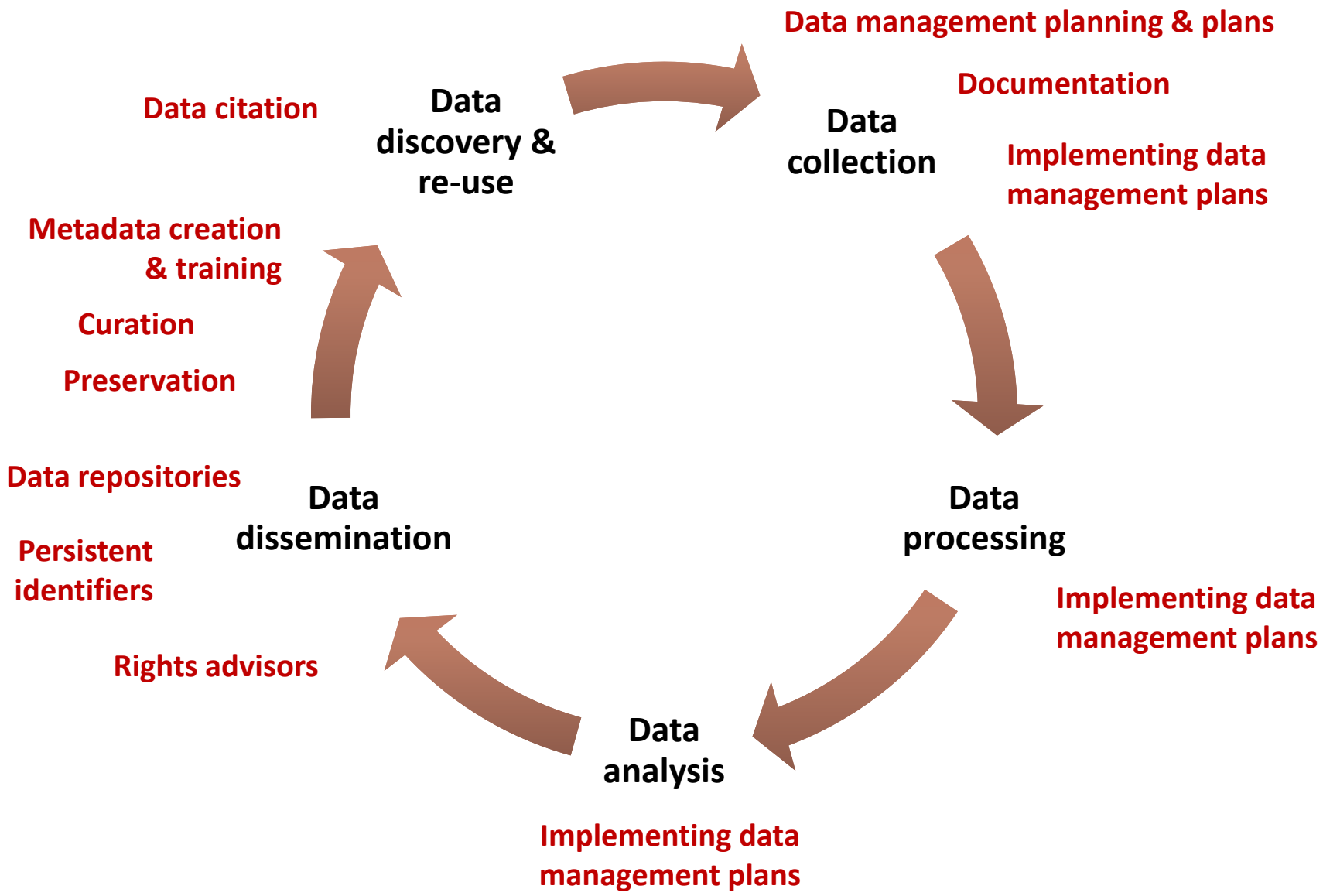
Research Life Cycle



Proposal development process



Data life cycle



New roles for a new vision

- Content
  - Data citation
  - Persistent identifiers
  - Documentation & metadata
- Services
  - Initiate conversations
  - Education & training
  - Finding data
  - Consultations
  - Advocacy
- Infrastructure
  - Institutional repositories, data repositories
  - Preservation networks
  - Data exchange networks
  - Support networks

BUILDING PARTNERSHIPS

A culture of success

- Strong liaison/subject librarian relationships with research faculty, staff, students, as well as programs and departments
- Trust in the library and librarians to preserve and curate their data
- Flexibility within the library to pursue new activities
- Possibility of failure accepted
- Recognized value of library as partner
- Mission and goals of library well-defined and related to institutional mission and goals

Building partnerships

1. Identify stakeholders
2. Figure out who is doing what
3. Identify the gaps, needs not being met
4. Meet the needs

Assessing the environment

1. Identify stakeholders
   - Funding data
   - Research Administration
2. Figure out who is doing what
   - Start with a conversation
3. Identify the gaps, needs not being met
   - Ask what they need
   - Ask for it in their words

Informal vs. formal strategies

Formal

- Focus groups
- Contextual interviews
- Survey questionnaires
- Electronic data & metrics

Informal

- Talk to your colleagues, subject librarians
- Talk to research faculty, staff, and graduate students
- Talk to administrators
- Talk to IT support for research-related units or services

Meeting the needs

- Identify your strengths and interest areas
- Start small, try out something new
- Consider sustainability - don’t overcommit
- Gather data to demonstrate success
- Learn from your mistakes
- Be flexible, stay flexible

Did we accomplish our goals?

- Increase awareness of the federal policies encouraging greater access to federally-funded data.
- Identify new roles and activities for librarians in data curation.
- Identify at what phases of the research process the library can integrate data and other services.

Selected resources

- National Academies of Science. (2009). Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. Retrieved from http://www.nap.edu/catalog.php?record_id=12615.
- Newton, M. P., Miller, C. C., & Stowell Bracke, M. (2010). *Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a Research Library Task Force*, Collection Management, 36(**1**), 53-67.
- Peters, C. & Riley Dryden, A. (2011). *Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston*, Science & Technology Libraries, 30(**4**), 387-403.
- Walters, T. & Skinner, K. (2011). *New Roles for New Times: Digital Curation for Preservation*. Washington, DC: Association of Research Libraries.

Thanks for your attention

Data Services Program

http://ulib.iupui.edu/digitalscholarship/dataservices

hcoates@iupui.edu

317-278-7125